DTIC FILE COPY

AD-A219 202

Putting it all together

Final comments

Allen Newell
Departments of Computer Science and Psychology
Carnegie Mellon University

Technical Report AIP - 61

The Artificial Intelligence and Psychology Project

Departments of
Computer Science and Psychology
Carnegie Mellon University

Learning Research and Development Center
University of Pittsburgh

90 03 12 021

2

Putting it all together

Final comments

Allen Newell
Departments of Computer Science and Psychology
Carnegie Mellon University

Technical Report AIP - 61

19 May 1988

In D. Klahr and K. Kotovsky (Eds.), *Complex Information Processing: The Impact of Herbert A. Simon.* Hillsdale, NJ: Erlbaum and Associates, 1988. (Twenty-First Carnegie Mellon Symposium on Cognition, Spring 1987)

This research was supported by the Computer Sciences Division, Office of Naval Research, under contract number N00014-86-K-0678. Reproduction in whole or in part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AIP - 61	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Carnegie-Mellon University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Computer Sciences Division Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213	7b ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Same as Monitoring Organization 8b. OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0678
8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS p4000ub201/7-4-86 PROGRAM ELEMENT NO: N/A; PROJECT NO: N/A; TASK NO.: N/A; WORK UNIT ACCESSION NO: N/A

11 TITLE (Include Security Classification)

Putting it all together: Final comments

12 PERSONAL AUTHOR(S) Allen Newell

13a. TYPE OF REPORT	13b. TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	88 May 19	46

16 SUPPLEMENTARY NOTATION In D. Klahr and K. Kotovsky (Eds.), Complex Information Processing: The Impact of Herbert A. Simon. Hillsdale, NJ: Erlbaum and Assoc., 1988.
(Twenty-First Carnegie Mellon Symposium on Cognition, Spring 1987)

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Herb Simon, unified theories of cognition, cognitive science, architectures; Soar

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

SEE REVERSE SIDE

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Alan L. Meyrowitz	22b TELEPHONE (Include Area Code) (202) 696-4302 22c. OFFICE SYMBOL N00014

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Abstract

There are many signs that cognitive science is moving significantly towards integrated treatments of its subject matter. The lifework of Herb Simon, in whose honor this symposium is being held, provides an important one. The papers of this symposium, towards whose critical appreciation this paper is supposed to contribute, provide another. And a recent extended reflection by the author on *Unified Theories of Cognition* (The William James Lectures at Harvard) provides the basis for yet a third. Against the background of an earlier reflection on the difficulties of cognitive psychology winning by playing the game of twenty questions with nature, these signs lead to a current upbeat assessment that cognitive science is putting it all together, and that it should continue to press forward.

PUTTING IT ALL TOGETHER
Final comments

Allen Newell

19 May 1988

Departments of Computer Science and Psychology
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

To be published in
D. Klahr and K. Kotovsky (Eds.), *Complex Information Processing: The Impact of Herbert A. Simon.* Hillsdale, NJ: Erlbaum and Associates, 1988.

(Twenty-First Carnegie-Mellon Symposium on Cognition, Spring 1987)

This research was sponsored by the Office of Naval Research (ONR), under contract number N00014-86-K-0678. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research.

List of Tables

PUTTING IT ALL TOGETHER
Final comments

I carefully chose my title — *Putting it all together* — to have multiple meanings. First, Herbert Simon has indeed put it all together over the course of his career. Thus, this is a suitable title for a volume that is a celebration of his cumulative work. But second, this title could be an echo of my William James Lectures: *Unified Theories of Cognition*, given this Spring of 1987 at Harvard. Thus, this is a suitable title to provide a lead in for me to write about what preoccupies me these days — always a good thing for a commentator to do. Third, the title could refer to putting the chapters of this volume all together. After all, I am officially a discussant — and the final one at that. So I could do what I was hired to do.

Three separate meanings might be thought to pose three alternatives, and thus a choice between them for the main line of a commentary. In fact, I believe they can be put together into a single exposition. For all three parts support the same point, namely, that our field is moving toward putting it all together. Indeed, I hope that the very act of my writing a commentary in this integrative mode will be seen to symbolize the need for us all to be synthetic — to put cognition together. However, paper being a linear medium, it is necessary to put the parts together, one after the other — first Herbert Simon, then William James, and then the volume. But that too can show that serial is as effective as parallel integration, and perhaps even more so.[1]

Herbert Simon

It might be thought that I am an especially good choice as commentator on Herb's career, having worked with him for so long. However, there is a difficulty. Herb had it all put together at least 40 years ago — and I've only known him for 35. The central idea is *bounded rationality* — there are limits on man as a decision maker and these limits, especially those of cognitive processing in all its varied forms, loom large in man's behavior. Everything that Simon has done has stemmed from the working out of this idea. This central scientific proposition has remained without revision.

A look, however superficial, through Simon's contributions to cognitive science, shows this clearly (Table 1). Each of these scientific topics — from the direct expression of a theory of bounded rationality, through symbol systems, to induction, chunking, task-acquisition, and onto spatial reasoning — all have to do with how humans deploy their limited processing capabilities so as to do their best with what they've got. That these parts of cognitive science have proven to have so much scientific substance, reflects how much our human actions are shaped by processing limitations.

---

[1]There is a whimsical tradition in computer science of self-referential acronyms. Thus *FINE*, an Emacs-like editor on Digital's Tops20 systems stands for *Fine Is Not Emacs*; and *Gnu*, which is a prefix for a family of systems such as *Gnu-Emacs*, stands for *Gnu is Not Unix*. In this tradition, the acronym for this paper is *PIAT*, which clearly stands for *PIAT Is All Together*. It is left as an exercise for the reader to find out how many multiple meanings there really are in this title.

Theory of Bounded Rationality

Adaptive systems (Simon, 1955a; Simon, 1956; Simon, 1980b)
Decomposable systems (Ando, Fisher & Simon, 1963)

Problem Solving

Search (Bayor & Simon, 1966; Simon & Kadane, 1975)
Human problem solving (Newell, Shaw & Simon, 1957; Newell & Simon, 1972)
Word problems (Paige & Simon, 1966)
Protocol analysis (Ericsson & Simon, 1984)

Symbol Systems

List processing (Newell & Shaw, 1957; Newell & Simon, 1976)
Architectures (Shaw, Newell, Simon & Ellis, 1959)
Semantic nets (Quillian, 1967)

Learning

Verbal learning and EPAM (Simon & Feigenbaum 1964; Simon & Gilmartin, 1973)
Adaptive production systems (Anzai & Simon, 1979; Neves, 1981)

Induction and Concept Formation

Patterns (Gregg & Simon, 1967; Simon & Kotovsky, 1963; Simon & Lea, 1974; Williams,1969)

Emotion (Simon, 1967)

Chunking

Size and rates of chunks (Simon, 1974; Gilmartin, Newell & Simon, 1976)
STM (Simon & Zhang, 1985)

Expertise :

Chess perception (Simon & Barenfeld, 1969; Chase & Simon, 1973, Simon & Gilmartin, 1973)
Physics (Larkin, McDermott, Simon & Simon, 1980)

Task Acquisition

Language (Coles, 1967; Simon & Siklossy, 1972)
Games, UNDERSTAND (Simon & Hayes, 1976; Williams, 1965)

Scientific Discovery

BACON, GLAUBER, STAHL, DALTON (Langley, Simon, Bradshaw & Zytkow, 1987)
KEKADA (Kulkarni & Simon, 1988)

Representation

Spatial reasoning and external memory (Larkin & Simon, 1987)

Table 1: Topics in cognitive science to which Simon and his colleagues have contributed (with representative citations).

It is an interesting side note that Herb did not succumb to the temptation of a capacity theory. A common response to limited processing is to posit a resource, call it *rationality juice*. A person has only a limited supply of this juice, and what is used for purpose P cannot be used for purpose Q. Then, the analyst regains the ability to apply optimization theory by assuming that the person will always distribute his limited rationality juice optimally among his options. Many have succumbed to positing such overall resource limits (by this or any other name). In my opinion, it has shielded them from discovering the real character of the mechanisms of cognition,

which have shape as well as volume. Instead, Herb went for the details of the specific mechanisms involved. There is nothing in Herb's story that I know of that says why this happened. But it is fortunate that it occurred, as the array in Table 1 bears witness.

I do not have to make the case at this point in the volume for how much Herb has put it all together. All of the chapters that have preceded me have done this job in greater or less detail. Even Herb's own chapter helps us see how all the pieces fit together. Still, I would like to add one example of my own. In 1975, in our joint talk accepting the ACM Turing Award, Herb and I chose to talk about *Symbols and Search* (Newell & Simon, 1972). We presented this as a *retrospective* account, not as a *prospective* scientific claim. The central role of search was clear by 1956 to Herb (and myself as his colleague), the central role of symbols by 1960. Indeed, we took the field to have understood these notions by the 1960s. That seems a long time ago, and it is. If anything constitutes the central dogma of cognitive science, it is these two ideas. They would seem to constitute the fountainhead of the subsequent research.

Note, however, that these two items are not at the top of Table 1. The topmost node of this generation tree of scientific knowledge is a *model of man*. The correct model of man as intendedly rational was already in place early on — indeed during the five years before I got to know Herb. Symbols and search are already the working out of this model, a refinement of it. Ed Feigenbaum is right, in his contribution to this volume, when he takes as key two very early papers of Simon that set out this model of bounded rationality (Simon, 1955a; Simon, 1956).

When I say that Herb has long since had it all put together, am I saying that Herb has known all these years all the science that the rest of us are still struggling to discover? Not at all. Being right in science does not mean knowing everything, or learning nothing new, or even not being surprised. Being right means being on the main path — the cumulative path. Along the main path, scientific revolutions occur in the metastructure of science or in the sociology, but not in the trenches. Consider the move from classical Newtonian mechanics to quantum theory. Since Kuhn (1962), all of us have learned to talk and think about this as a revolution. And indeed it was. But we must be careful to understand where the revolution occurred. It occurred in our overall views, in our big picture, in our heuristics. To be sure, there were technical developments — powerful and elegant ones. But they did not sweep the old away. Indeed, all of classical mechanics resides, alive and rosy well, within the new non-Newtonian view. If the French Revolution had been like this, it would have chosen the King to be the Minister of the Interior.

So I think Herb would agree with Ed Feigenbaum's comment that there has been a big shift towards knowledge intensive systems and towards understanding the powerful role played by having the right (or wrong) knowledge. I certainly would. But I doubt that Herb was surprised at the changed course of events. Each thing it is own time — bounded rationality, search, symbols, knowledge, architecture, learning, ... The reader must guess the next term in the sequence, for it is not predetermined, only constrained by a sort of scientific readiness. Indeed, the emergence of the focus on knowledge was not a surprise to me, but it surely was a major development. I rejoice that Ed and his colleagues at Stanford found the path and that it turned out to lead so far so fast.

Thus Herb, with a serenity and prescience that some of the rest of us lack, has always seen the field whole — has seen it as the unfolding of a single central idea. It has allowed him to move

from topic to topic within the area, always assured that the particular bit of the cathedral he happens to work on will add to the total structure — will help put it all together.

Unified Theories of Cognition

As I noted at the beginning, I have just finished a spring-full of lectures on unified theories of cognition (Newell, 1987). Let me introduce my concern with this topic, which relates strongly to putting it all together, by going back to a paper I gave at the Eighth Carnegie Symposium on Cognition (Newell, 1973a), organized by Bill Chase on visual information processing (Chase, 1973). The paper, itself a commentary just like this one, was entitled *You Can't Play 20 Questions with Nature and Win: Projective comments on the papers of this symposium.*

The situation, as it seemed to me in 1972, was that the cognitive view was in place and well taken. Recall that this was before the cognitive science movement of the late 1970s, although after Ulric Neisser's deservedly farr book *Cognitive Psychology* (1967). There was a great accumulation of experimental data, especially chronometric data. Indeed I was being called upon to comment on the papers of Mike Posner (1973), Lynne Cooper and Roger Shepard (1973), David Klahr (1973), and Bill Chase and Herb Simon (1973); and managed to squeeze in a reference to some other presenters as well (Bransford & Johnson, 1973; Clark, Carpenter & Just, 1973). This was as shining a collection of experimental luminaries as one could choose to read psychology by. Yet, despite the really fine examples of new data and penetrating analysis at that meeting, I feared for theoretical progress. It seemed to me, as I put it at the time, that the field did its theory by dichotomies, trying to find a general question to pose and then designing an experiment to settle *that* question, yes or no, and so move on down to the next subquestion. This is the classical strategy of twenty questions. And I did not think it would get us to the science of cognition that we all wanted.

I went on to describe some ways to try to put it together. One could create complete processing models. One could analyze complete tasks for all of the psychological phenomena they contain. One could take a single model and apply it to many different tasks across the psychological spectrum. In each case, at the heart of the enterprise is a cognitive model or detailed theory that provides the groundwork for an integration that is missing from the strategy of twenty questions.

Herb was pretty unhappy with my paper, though he didn't say much to me about it. But the depth of his feeling was evident when he entitled his own commentary in the Fifteenth Carnegie Symposium on Cognition (1979), *How to Win at Twenty Questions with Nature* (Simon, 1980a). Indeed, many people have reacted to my little paper, and reactions continue right up to the present day. (It is a bit disconcerting to find that it is one's commentary papers that seem to be the most read.) Most everyone (though not quite all) take it as pure criticism and even disillusionment — as showing that some commentators (to wit, me) believe that cognitive psychology is in deep trouble. Most everyone (again, not quite all) ignore or never notice that I also proposed three positive steps — that I was concerned with moving psychology toward a course that I judged would help put it all together. I even produced a companion paper (which was actually part of the same commentary, but split off for purposes of publication) that took a technical step toward architectures, namely *Production systems: Models of control structures* (Newell, 1973b).

Herb, I have no doubt, understood me, and in fact his essay seven years later bears that out. That essay reveals that he simply prefers a different metaphor. As he says, in concluding his discussion of the papers that were his assignment, "... three fine examples of bricklaying for a

cathedral that is beginning to take shape" (Simon, 1980a, p. 547). I have no trouble with the latter metaphor One of the great things about metaphors is that they can be put on and taken off with the weather, like so many sweaters. So it is easy to abandon the twenty-questions metaphor and put on the cathedral metaphor. In its terms, then, I was just trying to get the truckloads of bricks to go to the right construction site.

My goal for my commentary this time is to assert that it is coming together — or that the cathedral looks in good shape, rising majestically skyward. Or whatever other metaphor pleases you. The prognosis looks excellent to me and I have acted upon that assessment by undertaking my current scientific project, which is to move the field toward unified theories of cognition. To which we can now turn.

The Architecture as the Unified Theory

Let us start with a familiar notion — the architecture — sketched out in Figure 1. An architecture is the fixed structure that realizes a symbol system. It is what makes possible the hardware-software distinction — a structure that can be programmed with content (encoded in some representation), which can itself control behavior. As the figure shows, intelligent systems are constructed in levels. The two levels of main interest for cognitive science are the symbol level and above it the knowledge level (Newell, 1982). The knowledge level abstracts away from all representation and processing, to reflect only what the system has acquired about its environment that permits it to deploy its resources to attain its goals. Knowledge-level systems are just a way of describing a real system, of course. Thus, such a system is also always describable in more physical terms — with the representation and processing accounted for. Such systems are symbol-level systems (Newell, 1980).

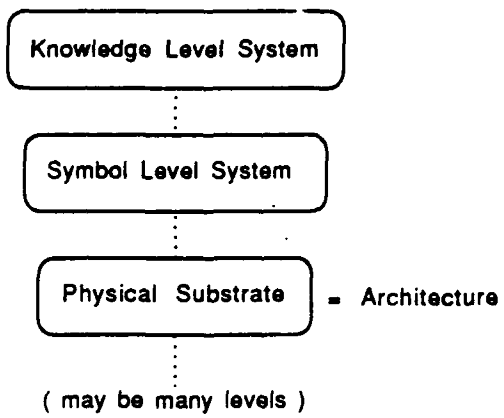


Figure 1: The architecture and system levels.

The symbol level is that of data structures with symbolic operations on them, being carried out under the guidance of plans, programs, procedures or methods. These are organizations of processing that are rationally ordered to the attainment of the goals of the system — that do the work that makes it true that the system uses whatever it knows to attain its goals. A symbol system, also, is just a way of describing a real system, so that this same system is also anlways describable in terms of some physical substrate.

This substrate (or rather, its design) is the architecture. In current digital computers it is a

register-transfer level system, but in biological systems it is some organization of neural circuits. As George Miller said in his chapter, the separation of hardware from software was Simon's essential simplification — that it was at the heart of the conceptual position Simon had taken. George is absolutely right. To say that a system has an architecture is another way of saying that it admits of being programmed, which is to say that it admits of the hardware-software distinction.

Unified theories of cognition are architectures. That is a key point. It is the architecture, changing only slowly over time, that provides the communality that shows up amid the diversity of behavior of a single human. It is the communality of the architecture across all humans that makes the behavior of all humans similar in many ways. To use the slogan of this commentary: It is the architecture that keeps it all together. Thus, to propose a unified theory of cognition is to propose a cognitive architecture.

This volume sports a veritable showcase of cognitive architectures — enough so I've listed them in Table 2. First there is CAPS, the cognitive architecture that lies behind the work on working memory in the context of reading, discussed by Marcel Just and Pat Carpenter (Thibadeau, Just & Carpenter, 1982). Next, there are various unadorned productions systems. Neil Charness used an informal production system for his hand simulations; Jill Larkin used a version of Ops5 (ExperOps5). John Anderson's work on errors was built entirely around his use of PUPS, a production-system based architecture that is the high-level successor to Act* (Anderson, 1983). BRIAN, the topic of Steve Kosslyn's chapter is special in a couple of ways. Not only is it a neural architecture, but the focus of his chapter is *to establish* BRIAN as a viable candidate architecture. The other speakers *use* their architectures in explicating whatever scientific story they tell. Finally there is Soar, the architecture by John Laird, Paul Rosenbloom and myself (Laird, Newell & Rosenbloom, 1987). Its inclusion in the list provides a living example of a *writing act* — the written analogue of a speech act. The list contains the cognitive architectures discussed in this volume, and Soar joins that group by virtue of my having added its name to that list, and discussing the addition as a writing act.

CAPS (Just, Carpenter & Thiebadeau)

Unadorned Production Systems

Informal hand simulation (Charness)
ExperOps5 (Larkin)

PUPS (Anderson)

BRIAN (Kosslyn, Sokolov, Flynn & Chen)

Soar (Laird, Newell & Rosenbloom)

Table 2: Specific symbolic architectures in residence in this volume.

This list indicates the shift in cognitive science towards integrated theoretical attempts and away from dichotomies. These authors seek *systems of mechanisms* that explain the behavior of subjects on their tasks. In general they use their systems for explanation, rather than just propose or analyze them. The milage obtained by using an architecture seems to me quite evident. This

is certainly what I had in mind fifteen years ago. This represents real progress. I congratulate the authors of these chapters, one and all. They themselves will have to answer whether my twenty-questions carp helped nudge them in this direction, so that essay can claim to have played a positive role in the evolution of our science, rather than just the negative, critical role that so many commentators have assigned to it. Actually, I care much less about that than about the fact of the case, namely, that cognitive science is outgrowing the twenty-questions games of its youth.

However, I am left with some questions, though these are mostly to the other participants. To Kevin Dunbar and David Klahr , and also to Anders Ericsson and Jim Staszewski — Why not use an architecture as the basis for detailed processing models? It seems to me the complexity of both your tasks cries out for it. It would help to see through all of that complexity. Likewise, to Ken Kotovsky and David Fallside — Why not use an architecture? It would make lots of difference to the questions you ask. Finally, to Marcel Just and Pat Carpenter — Why not make more use of CAPS? It sits behind your models already, and I think it could play a much more explicit and important role.

Three factors modulate this assertion of the centrality of architecture to theories of cognition. First, in adaptive systems, the nature of the task has a strong determining effect on behavior (Simon, 1980b). Therefore, in so far as humans do similar tasks (seek similar goals in similar environments), they behave similarly. Second, the knowledge available determines how an intelligent system behaves. Therefore, in so far as humans have the same knowledge in doing the same tasks, they behave similarly. Third, the knowledge they have is determined by their prior experiences, including those we call education and socialization. Therefore, in so far as humans are educated and socialized similarly, they will behave similarly.

At their strongest, these modulations imply that systems are describable at the knowledge level, so that their only communalities are those of knowledge, goals and the surrounding constraining environment. This is the important plank in Ed Feigenbaum's general research stance, which he has made the central substantive theme of his chapter in this volume. Much vanishes at the knowledge level — most of we think of as psychology and also the architecture. More precisely, the architecture still has a job to do, but it never shows its face. Now, in fact, the architecture is very much in evidence (and with it psychology). The modulations do not render it imperceptible. There is more here than meets Ed Feigenbaum's eye.

Herb has championed a particular middle ground here, namely that a small number of parameters define the effects of architecture and thus express the ways in which human rationality is bounded even beyond the limits of knowledge that stem from social, educational and organizational locality. Primary among these are basic limits on the speed and concurrency of processing, short-term memory, and the acquisition of new long-term knowledge (Newell & Simon, 1972; Simon, 1979a, Parts 1 and 2). Such a view focusses on the parameters and keeps the rest of the architectural structure essentially in the background. As Herb has shown in many analyses, there is substantial power in this approximation.

My own view at the moment is that more architectural detail will be found relevant to a useful model of human behavior. In this regard, Table 3 shows a favorite list of mine — all the constraints that jointly bear on the human architecture to make it what it is. The dimension of flexibility, including symbolic, abstract and linguistic abilities, is what computation provides. It

is what leads on to the notions of universal computation, knowledge is all, and ultimately that the architecture is immaterial, merely the turtle on which the towers of elephants stand that hold up the world — so far down as to be consignable to mythology.

1. Behave flexibly as a function of the environment
2. Exibit adaptive (rational, goal-oriented) behavior
3. Operate in real time
4. Operate in a rich, complex, detailed environment
   - 4.1 Perceive an immense amount of changing detail
   - 4.2 Use vast amounts of knowledge
   - 4.3 Control a motor system of many degrees of freedom
5. Use symbols and abstractions
6. Use language, both natural and artificial
7. Learn from the environment
8. Acquire capabilities through development
9. Live autonomously within a social community
10. Exhibit awareness and a sense of self
11. Be realizable as a neural system
12. Arise through evolution

Table 3: Multiple constraints on the nature of the architecture.

But flexibility and its associated constraints are only part of the story. Other constraints also play a strong role in shaping human cognition — to behave in real time, to learn and develop continuously, and to be realizable by evolutionary processes operating on a neural technology, built on a genetic substrate. The architecture reflects these requirements as well. In doing so, it affects the way flexibility is realized and its limits. These constraints shape the architecture in specific ways that may show in behavior.

Let me focus on just one of these — the temporal constraint. The time scale at which things happen is a critical characteristic. Consider Table 4, another favorite of mine, the time scale of human action. It also shows the systems hierarchy. Each level is a quite different system, starting from organelles at the bottom of the figure, up through neural circuits to the cognitively behaving individual and on up to social systems and beyond. This is a true systems hierarchy, in which each system is composed of the components of the level below — neurons out of organelles, neural circuits out of neurons, etc. Each level is larger in size and runs slower than the level below, an inevitable consequence of having the level below as components. Interestingly, each system level is only about a factor of ten larger and slower than its components. This is just about the *minimum* factor that is required for building up a new system — it is necessary to have a few components to interact and to give them a few operation times to pass the interactions around, before new behaviors emerge. That the systems levels pile up about as fast as possible should come as no surprise to most of us here, since it is (in part) a reflection of considerations Herb put forth so cogently twenty-five years ago in his *Architecture of complexity* (Simon, 1962). The systems that we see around us are hierarchical, because

hierarchies are stable and because unstable systems do not survive. Delving into the argument a bit shows that ceteris paribus the smaller the systems level, the more stable.

TIMESCALE OF HUMAN ACTION

Scale (secs)	Time Units	System	World (theory)
10^7	months		
10^6	weeks		SOCIAL BAND
10^5	days		
10^4	hours	Task	
10^3	10 mins	Task	RATIONAL BAND
10^2	minutes	Task	
10^1	10 sec	Unit task	
10^0	1 sec	Operations	COGNITIVE BAND
10^{-1}	100 ms	Deliberate act	
10^{-2}	10 ms	Neural circuit	
10^{-3}	1 ms	Neuron	NEURAL BAND
10^{-4}	100 µs	Organelle	

Table 4: Timescale of human action.

I want to call your attention to the *cognitive band*, starting above neural circuits, which are at ~10 ms, and moving up to the first observable cognitive behavior, at about 1 s. The *real-time constraint on cognition* is that the human must produce genuine cognitive behavior in ~1 s, out of components that have ~10 ms operation times. This is only about 100 operation times to get from the basic circuitry to behavior. Put in terms of system levels, 100 operation times is only two levels, each of 10 operations (that is, 100 = 10x10 = 10 operations of components, each of which itself takes 10 operations of its components). To be explicit, there is *hardly any time at all* to produce complex behavior. This constraint has been widely noted. The connectionists in particular have used it as an argument for why the architecture has to be massively parallel (Feldman & Ballard, 1982) — and why clunky old serial symbolic architectures are simply out of the race. However, much more follows from it than the need for parallelism. It is binding enough to shape many aspects of the cognitive architecture.

Figure 2 shows the levels of the cognitive band whose existence and structure can be inferred from the real-time constraint on cognition, given the neural-circuit level on the low end and the representational and computational requirements on the high end (Newell, 1986; Newell, 1987, Lecture 3). Let me take just a sentence apiece to indicate these levels, without going through the argument. The bottom level of the cognitive band provides *symbolic access*. The next level up provides the most *elementary deliberation*, comprising multiple accesses of memory to accumulate the considerations that enter into a deliberate selection of an action. The next level provides *simple operations*, composed of sequences of deliberations — simple, because the operations themselves must be immediately realizable, hence pre-existing (although selected from an available repertoire). The top level of the cognitive band admits composing a response or result by working with operations which themselves are composed and hence can be adapted to the task at hand. This is the first level that admits full problem spaces, i.e., spaces with operators adapted to the task at hand.

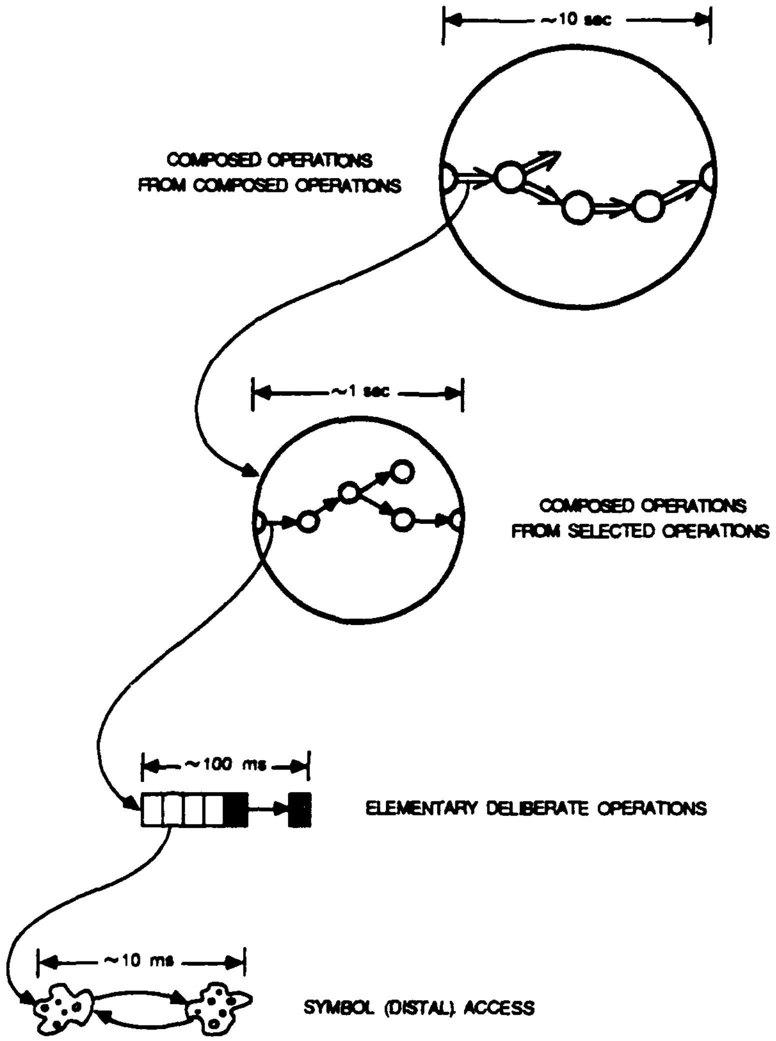


Figure 2: Levels of the cognitive band.

As the system builds on up from this top cognitive level, employing ever more complex operations, we begin to move into the world that Ed Feigenbaum champions — the world of *intendedly rational* behavior. Increasingly there is enough computational space to permit the knowledge available in the system to be the strongest determinant of the system's response. Adaptation has the opportunity to be effective. The structure of the behaving system is dictated not by the architecture but by the structure of the task as perceived by the agent. It becomes a different world than what the cognitive psychologist normally studies. However, as Herb repeatedly emphasizes, it never becomes quite what the economists wish for, where *only* the objective external economic situation counts. It remains only intendedly rational, subject to the limits of knowledge and computing power.

But back at the level of cognitive machinery, it is not this way. Indeed, it will be observed that the system arrives at the level of immediate action (~1 s) before it is able to deploy other than simple pre-existing operations to compose its response. The real-time constraint on cognition does not provide time for more.[2]

Let me note some of the consequences of Figure 2. A chief one is that the time-mapping from cognition to neural systems is fixed in terms of orders of magnitude. Symbol access takes ~10 ms. It could not take ~100 ms, because there would never be time for cognitive behavior at ~1 s. Elementary deliberation occurs at ~100 ms. This is as soon as it can occur, given ~10 ms for symbolic access, and as late as it can occur, given the requirement for behavior at ~1 s. Simple operations occur at ~1 s, and immediate reactive behavior must result from such processes.

The significance of such a mapping, however approximate, should not be underestimated. For years, cognitive psychology has enjoyed the luxury of considering its analysis to be one that floats entirely with respect to how it might be realized in the brain. This has been both a reflection of the long standing chasm between brain and behavior and a contributor to it. Figure 2 signifies that era is coming to an end. Of course, the era is not ending *because* of the figure. A look at the architecture discussed by Steve Kosslyn shows how much the neurophysiological and the behavioral are converging. However, I believe that the structure of Figure 2 goes a long ways towards seeing how we must map results from neural and cognitive descriptions into each other. The floating kingdom has finally been grounded.

The yield of Figure 2 is not limited to just the temporal mapping. From it, plus the considerations behind it, comes the basic recognitional character of the symbolic level, the existence of automaticity at the ~100 ms level, problem spaces, and the existence of a process that continually converts experience into recognitions. The arguments are of course speculative. But, importantly, they are not tied to any specific architecture. Rather, they serve to provide constraints that the human cognitive architecture must satisfy, and thus help to guide our search for it.

---

[2]The compulsive force behind the constraint — why overt behavior *must* occur by ~1 s rather than ~10 s or ~100 s — would appear to be evolutionary. If organisms can respond this fast, given their technology (here, neural technology), then they must. For whoever gets there fastest survives.

Soar: A Candidate Cognitive Architecture

Figure 2 sets the stage for a brief description of Soar, as a candidate theory of the human cognitive architecture. Soar has been around for some time as an AI architecture for general problem solving and learning (Laird, Newell & Rosenbloom, 1987; Laird, Rosenbloom, Newell, 1986). Soar is based on many of the mechanisms that have played a major role in information processing theories, such as problem spaces and production systems. Thus, from the start, Soar has been an architecture that is roughly commensurate with what we know of human cognition. But further development and analysis have convinced us that Soar is a serious candidate for the architecture of human cognition in detail as well as in over-all character (Newell, 1987). Soar is an architecture for putting it all together.

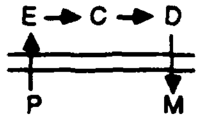
Figure 3 enumerates the main mechanisms in Soar. The top four items set the outer context, being aspects shared by all comprehensive cognitive-science theories of human cognition. Soar operates as a controller of the human organism, hence is a complete system with perception, cognition and motor components. This already takes mind in essentially functional terms — as the system that arose to control the gross movements of a mammal in a mammalian world. Soar is goal-oriented with knowledge of the world, which it uses to attain its goal. This knowledge is represented by a symbol system, which means that computation is used to create representations, extract their implications for action, and implement the chosen actions. Thus, Soar is an architecture, with most of the knowledge in the total system embodied in the content that the architecture makes meaningful and accessible.

The rest of the items describe Soar from the bottom up, temporally speaking. Soar comprises a large *recognition memory*. This is realized by an Ops5-like production system (Forgy, 1981), with a set of productions each of whose conditions is matched against the elements in working memory, leading to the execution of the actions of the successful instantiations. The productions execute very rapidly, within about 10 ms. Although in AI and cognitive science, productions are usually taken to correspond to operators (deliberately deployed actions), here they correspond to associational memory. Thus, production actions behave like a memory retrieval. They only enter new elements into working memory; they cannot modify or delete what is there; and there is no conflict resolution (of the kind familiar from Ops5). Each production operates independently — an isolated memory access and retrieval.

The next level of organization, which occurs within ~100 ms, consists of the *decision cycle*. This comprises a sequence of retrievals from long term memory (i.e., a sequence of production firings) that assemble from memory what is immediately accessible and relevant to the current decision context. This sequence ultimately terminates, when no more knowledge is forthcoming (in practice it quiesces quickly). Then a *decision procedure* makes a choice of the next step to be taken. This changes the decision context, so that the cycle can repeat to make the next decision. At the 100 ms level, cognitive life is an endless sequence of assembling the available knowledge and using it to make the next deliberate choice.

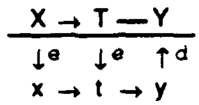
The decisions taken at the 100 ms level implement search in *problem spaces*, which comprise the next level of organization, namely, at the ~1 sec level. Soar organizes all its goal-oriented activity in problem spaces, from the most problematical to the most routine. It performs a task by creating a space within which the attainment of the task can be defined as reaching some state, and where the moves in the space are the operations that are appropriate to performing the

1. Controller - Perception-Cognition-Motor



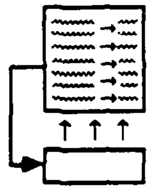
2. Knowledge and Goals

3. Representation, Computation, Symbols



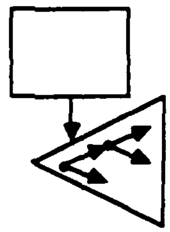
4. An architecture plus content

5. Recognition memory (about 10 ms)



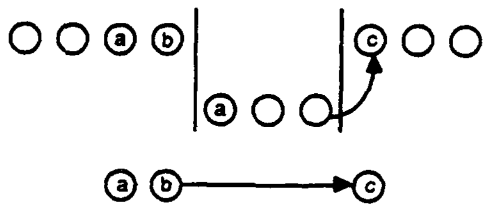
6. Decision cycles - Automatic (about 100 ms)

7. Problem spaces and operators (about 1 sec)



8. Impasses and Subgoals

9. Chunking (about 10 s)



10. Intended rationality (100 sec and up)

Figure 3: Soar as a unified theory of cognition.

task. The problem then becomes which operators to apply and in what order to reach the desired state. The search in the problem space is governed by the knowledge in the recognition memory. If Soar has the appropriate knowledge and if it can be brought to bear when needed, then Soar can just put one operator in front of another, to step its way directly to task attainment. If the memory contains little relevant knowledge or it can't be accessed, then Soar must search the problem space, leading to the familiar combinatorial explosion.

Given that the problem-space organization is built into the architecture, the decisions to be made at any point are always the same — what problem space to work in; what state to use (if more than one is available); and what operator to apply to this state to get a new state, on the way to a desired state. Making these choices is the continual business of the decision-cycle. Operators must actually be applied, of course — life is not all decision making. But applying operators is just another task, which therefore occurs by going into another problem space to accomplish the implementation. The recursion bottoms out when an operator becomes simple enough to be accomplished within a single decision cycle, by a few memory retrievals.

The decision procedure that actually makes the choice at each point is a simple, uniform process that can only use whatever knowledge has accumulated via the repeated memory searches. Some of this knowledge is in the form of *preferences* about what to choose — that one operator is preferred to another, that a state is acceptable, that another state is to be rejected. The decision procedure takes whatever preferences are available and extracts from them the decision. It adds no knowledge of its own.

There is no magic in the decision cycle — it can extract from the memory only what knowledge is there, and it may not even get it all; and the decision procedure can select only from the options thereby produced and by using the preferences thereby obtained. Sometimes this is sufficient and Soar proceeds to move through its given space. Sometimes — often, as it turns out — the knowledge is insufficient or conflicting. Then the architecture is unable to continue — it arrives at an *impasse*. This is like a standard computer trying to divide by zero. Except that, instead of aborting, the architecture sets up a *subgoal* to resolve the impasse. For example, if several operators have been proposed but there is insufficient information to select one, then a *tie impasse* occurs, and Soar sets up a subgoal to obtain the knowledge to resolve the tie, so it can then continue.

Impasses are central to Soar. They drive all of Soar's problem solving. Soar simply attempts to execute its top level operators. If this can be done, Soar has attained what it wanted to do. Failures along the way imply impasses. Resolving these impasses, which occurs in other problem spaces, can lead to other impasses, hence to subproblem spaces, and so on. The entire subgoal hierarchy is generated by Soar itself, in response to its inability at performance time to attain its objectives. The different types of impasses generate the full variety of goal-driven behavior familiar in AI systems — operator implementation, operator instantiation, operator selection, precondition satisfaction, state rejection, etc.

In addition to problem solving, Soar learns continuously from its experience. The mechanism is called *chunking*. Every time Soar encounters and resolves an impasse, it creates a new production (the chunk) to capture and retain that experience. If the situation ever recurs, the chunk will fire, making available the information that was missing on the first occasion. Thus, Soar will not even encounter an impasse on a second pass.

The little diagram at the right of *chunking* at the bottom of Figure 3 sketches how this happens. The view is looking down on working memory, with time running from left to right. Each little circle is a data element that encodes some information about the task. Starting at the left, Soar is chugging along, with productions putting in new elements and the decision procedure determining which next steps to take. At the vertical line an impasse occurs. This sets a new context and then behavior continues. Finally, Soar produces an element that resolves the

impasse (the circled c at the second vertical line), after which behavior continues in the original context. The chunk is then built, with an action corresponding to the element c that resolved the impasse and with conditions corresponding to the elements prior to the impasse that led to the resolution (the circled a and b). This captures the result of the problem solving to resolve the impasse, and does so in a way that permits it to be evoked again to avoid that particular impasse.

Chunking operates as an automatic mechanism that continually caches all of Soar's goal-oriented experience, without detailed interpretation or analysis. As described, it appears to be simply a practice mechanism — a way to avoid redoing the problem solving to resolve prior impasses, thus speeding up Soar's performance. However, the conditions of the productions reflect only a few of the elements in working memory at the time of the impasse. Thus, chunks *abstract* from the situation of occurrence, and can apply in different situations, as long as the specific conditions apply. This provides a form of *transfer* of learning. Although far from obvious, this mechanism in fact generates a wide variety of kinds of learning (Steier, Laird, Newell, Rosenbloom, Flynn, Golding, Polk, Shivers, Unruh & Yost, 1987), enough to conjecture that chunking might be the only learning mechanism Soar needs. Chunks get built in response to solving problems (i.e., resolving impasses). Hence, they correspond to activities at about the 1 sec level and above. The chunk, of course, is a production, which is an entity down at the memory access level at about 10 ms.

The higher organization of cognitive activity arises from top-level operators not being implementable immediately with the information at hand. Thus, they must be implemented in subspaces with their own operators, which themselves may require further subspaces. Each descent into another layer of subspaces means that the top-level operators take longer to complete — i.e., are higher level. Thus the timescale of organized cognitive activity climbs above what can be called the region of cognitive mechanism and into the region of intendedly rational behavior. Here, enough time is available for the system to do substantial problem solving and use more and more of its knowledge. Here, the organization of cognition becomes increasingly dictated by the nature of the task and the knowledge available, rather than by the structure of the architecture.

There are limits to the extent of the means-ends structure that a human builds up at a given moment, in response to impasses. The really long-term organization of behavior cannot arise just from piling up the goal hierarchy. Soar does not yet incorporate any theory of what happens as the hours grow, disparate activities punctuate one another, and sleep intervenes to let the world of cognition start afresh each morning.

What can be gleaned from such a rapid-fire tour through Soar? Certainly not an assessment of whether it is an effective theory of human cognition. However, perhaps it can be glimpsed that Soar is an architecture that spans the entire range of psychological functions — well, not quite the entire range yet, but it is reaching in that direction, with its modeling of behavior from the ~10 ms level on up to into the rational band of ~1000 s, namely, the time it takes humans to solve problems such as designing algorithms (Steier & Newell, 1988). This lets me make my point — that cognitive science is on the threshold of obtaining architectures that will provide the basis for comprehensive theories. This is a major step toward being able to put it all together.

The Chapters of this Volume

Having made my point about the central integrative role of architectures, let me finally turn to the chapters of this volume. My job, of course, is to put them all together. Actually, I will avoid my obligation directly, by turning the point around and asking what these chapters tell us about putting it together. Thus, I will keep to my central theme. But as a consequence, I will not criticize the chapters as studies in their own terms.

It is not a single activity to put a science together. Cognitive science, like any science, has multiple aspects: methodology, foundations, the heartland, and the borders. The chapters of this volume say something about putting matters together in each of these aspects.

Methodologies

Cognitive psychology has developed many methodologies for studying human behavior. It inherits the basic methodologies of controlled experimentation, statistical design and data analysis from its ancestral psychological tradition. But it has also been prolific in creating new methodologies and sharpening existing ones. Table 5 lists a number of examples. The first four — task analysis, mental chronometry, simulation and protocol analysis — are solidly in place. They all go back to the first decade of the cognitive revolution. Actually, I like George Miller's remark in his chapter that the behaviorists were the revolutionaries, so the events commencing in the 1950s should be taken as the counter-revolution. I have so labeled the figure, for the counter is nowhere more apparent than in the methodologies that were used. Interestingly, protocol analysis, perhaps the epitome of the countermove, has taken a long time to become widely practiced. The last six methodologies are also familiar, but their use is much more specialized and scattered. Indeed, the fifth one — theorizing within a theoretically specified architecture — has hardly begun. As I have just borne witness, it is my own current project to try to help it along.

1.	[TA]	Task analysis (including AI systems)
2.	[RT]	Mental chronometry
3.	[Sim]	Simulation
4.	[PA]	Protocol analysis
5.	[Arch]	Architecture
6.	[SS]	Special subjects: neurological deficits, experts
7.	[CA]	Comparative analysis: Novice / Expert, Child / Adult
8.	[EM]	Eye movements
9.	[QEA]	Qualitative error analysis
10.	[ET]	Experimental training

Table 5: Methodologies of the cognitive counter-revolution.

Methodologies partition a field as surely as theoretical ideas. Indeed, the tensions within cognitive science have on occasion been laid to the different methodologies of its subdisciplines — psychology with its experimental subjects, linguistics with its individual informants,

philosophy with its imagined situations and artificial intelligence with its designed programs. Within psychology, topics become identified with particular methodologies and develop their own special groups of investigators. One need only note the way error measures and a few experimental paradigms dominated verbal learning for years, or the specialized territory occupied by eye movement research over the years, with its own special conferences and books (Fisher, Monty & Senders, 1981; Groner, Menz, Fisher & Monty, 1983; Monty & Senders, 1976; Senders, Fisher & Monty, 1978).

Table 6 shows the methodolgies used by the research reported in the volume. Of course, they use a wide variety of the methods; that is only to be expected. More interesting, there is a strong tendency towards the use of multiple methodologies, substantially more than two. I take this as a sign of integration — less purity and more power. It is an interesting step toward putting it together to be able to relate the observations from many different sources. One is tempted to reach for some analogy to the use of multiple knowledge sources in AI, as a mark of intelligence. But the causal arrow probably goes the other way.

Kosslyn	Sim, Arch, SS
Just-Carpenter	RT, PA, Arch, EM
Kotovsky-Fallside	TA, RT
Dunbar-Klahr	TA, PA, CA
Feigenbaum	TA (Foundations)
Charness	TA, PA, SS, Arch
Hayes	PA, CA
Ericsson-Staszewski	RT, PA, SS, ET
Greeno	(Foundations), PA
Larkin	TA, Sim, Arch, QAE
Anderson	TA, Sim, Arch, QAE
Simon	(Theory applications), SS

Table 6: Multiple methodologies in the studies of this conference.

Foundations

Foundational issues were addressed by both Ed Feigenbaum and Jim Greeno. I have strong opinions about what each said — but that is a hallmark of foundational issues. One issue from each seems relevant to putting cognitive science together.

Preparation vs deliberation

Although I agree with much of what Ed says in his chapter, I wish to differ with him on what he calls the knowledge versus search display. I present my version of it in Figure 4. The axes are labeled *preparation* along the vertical, and *deliberation* along the horizontal. The diagram refers to the means by which a system performs a task. Preparation is the extent to which the system draws on what it has prepared in advance of the task. Deliberation is the extent to which the system processes information once the task is set — engages in searching problem spaces, or reasons from what it knows, or whatever you want to call it. The curves represent equal-

performance isobars. That is, different choices of how much to draw on prepared material and how to compute once the task is set can yield the same performance — more preparation and less deliberation versus less preparation and more deliberation.

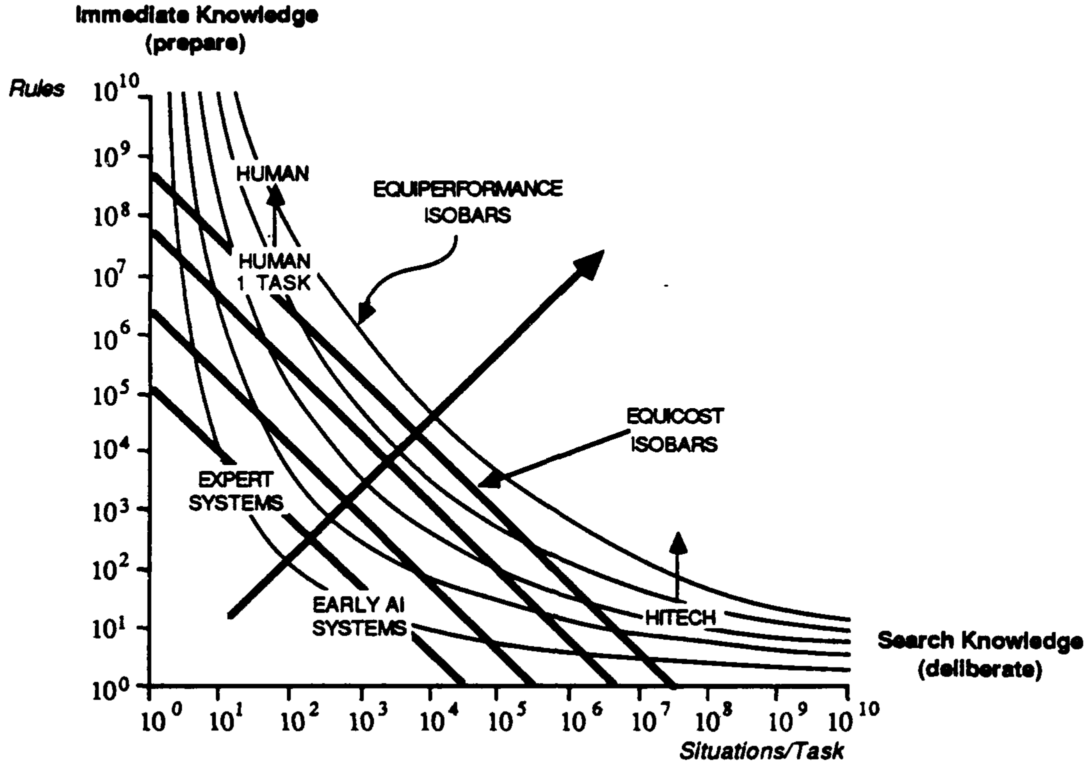


Figure 4: Preparedness vs deliberation tradeoff.

This graph is a variant of the familiar store-versus-compute tradeoff (Berliner & Ebeling, 1988). It is also often called the knowledge vs search graph, which is the term Ed uses. But the latter phrase is something a misnomer. Both axes represent knowledge — knowledge that comes from stored memory structures (what has been prepared) and knowledge that comes from computation during the decision process. The axes, however, are not measured in knowledge abstractly. Stored knowledge is measured in amount of structure (number of rules or number of memory bits) and acquired knowledge is measured in situations examined or processed.

This tradeoff is fundamental to information processing systems. Different systems embody different strategies and end up in different places on the diagram. Although the graph refers to the division used for a given task, systems typically treat all their tasks similarly. Thus, a system itself can be located at a point in the space in the middle of the cluster of its task points. Thus, Figure 4 shows the characteristics of various types of AI systems — the early AI search-oriented systems, which had small knowledge and modest search, expert systems which have more knowledge (up to ~10^4 rules currently) but do less search. Hitech, Hans Berliner's high-master chess machine (Berliner & Ebeling, 1988) is way out at the extreme of deliberation, with ~10^7 situations examined per external move, and with only a small amount of recognitional knowledge.

This diagram is part of the foundations of AI and cognitive science. It provides a fundamental view of how information processing systems can differ and yet be related to each other. It tells us that systems, such as Hitech, are not an entirely unrelated way of attaining task performance (to be classified as brute force), but rather a different part of the total space of information processing systems. This is a space of architectures we need to explore in understanding intelligence. Exactly the same considerations enter into systems such as Hitech as into other systems in the space. For instance, once the architecture of a system is fixed, whether for a human or Hitech, the amount of deliberation becomes fixed. Then the system can improve only by vertical movement in the space. Indeed, Hitech has moved from a low expert to a high master entirely by adding recognition knowledge. As noted above, the total amount of recognition knowledge involved is quite small — it is as if the hyperbolic character of the diagram really applied and the isobars are squeezed tightly together as one moves out toward asymptotically high search.

As another instance, it is possible to have systems that move back and up along an isobar, i.e., decreasing deliberation and increasing preparation. Soar, with its chunking, is such a system. It appears much more difficult to move in the other direction. In fact, I do not know any systems that do so in any substantial way. This seems to be the crux of Ed's criticism — that a Mycin cannot extend search and do with less knowledge. Indeed, that is true. In the short run we run out of knowledge just as we run out of time. And we can only search the spaces we know about (more knowledge). Time scale, as always, is crucial. Thus, the just-noted continual movement by chunking from search to knowlege does not occur within a single performance, but takes many trials. Furthermore, new knowledge can be generated by extended thought if the spaces are available that can support it. That Mycin does not have such spaces — that, in the parlance of expert systems, it is a *shallow* system, not a *deep* one — only reveals that Mycin and most early-generation expert systems, were primarily explorations in what could be attained with only stored knowledge. They are no more to be taken as the shape of a generally intelligent system than are the early AI search systems. Fully intelligent systems will do extended search to add to their knowledge, just as mathematicians do in searching for proofs.

Sufficiency of physical symbol systems

I also agree with much of what Jim Greeno says in his chapter. In particular, I think there is much to be learned about how humans deal with external environments and how they use the environment to keep track of what they are doing and to perform computations for them, both implicitly and explicitly. There are many signs that more attention is being paid to such things. One is the strong emphasis placed on situated action, such as by the Center for the Study of Language and Information at Stanford and nearby West Coast research centers. At the present conference, Jill Larkin's analysis of coffee making fits within this focus rather explicitly. Topics become ripe for exploration in particular epochs. I agree that this one seems ready now, and I hope the immediate future will see a major increase in our understanding of how cognitive agents work in intimate concert with the world.

But Greeno goes somewhat further than that, and grounds this shift in a need for a new philosophical view of symbols and how they refer to the external world. In that, he seems to me to be wrong, though undoubtedly not wrong that some scientists perceive it that way. I fail to see that the current conceptual apparatus is inadequate for dealing with situated action and close-coupled interaction with the outside world, at least of the kind that Greeno is discussing. This is

the aspect of Greeno's chapter that is relevant to my theme of putting it all together. Arguments for shifts in the foundations are often meant to signal, not a way of putting it together, but a way of producing a new start. Greeno, good scientist that he is, specifically acknowledges the way the new future grows out of the past. Still, I wish to quarrel a bit.

To be specific, the concept of symbols that has developed in computer science and AI over the years is not inadequate in any way that I understand. It does not need extending in some special way to deal with the external world. It is not especially inward looking. Symbols, as that concept occurs in physical symbol systems (Newell, 1980), designate entities in an external world, including actions to be taken to effect changes out there. No implicit notions of context freedom exist to plague the formulation, so as to pose difficulties to being indexical or relative or operating off concurrently perceived external structure.

At first blush, I cannot imagine how one would think otherwise. For example, such symbols are used as a matter of course by the Navlab autonomous land vehicle (a van that drives itself around Schenley Park next to Carnegie-Mellon), which views the road in front of it through TV eyes and sonar ears, and controls the wheels and speed of the vehicle to navigate along the road and between the trees (Thorp, Herbert, Kanade & Shafer, 1987). The symbols that float everywhere through the computational innards of this system refer to the road, grass and trees in an epistemologically adequate, though sometimes empirically inadequate, fashion. These symbols are the symbols of the physical symbol system hypothesis (Newell & Simon, 1976), pure and simple.

Well, maybe I can imagine how one might think otherwise. Here is one possibility. The mechanics of symbols are built around the *access* relation, which permits the system, upon encountering a symbol token in the course of processing a symbolic expression, to access the additional symbol structures that are related to the meaning of the tokened symbol. Those symbol structures are still internal to the computational system. These structures will, in general, contain additional symbol tokens leading to further access to other symbol structures. Around and around it goes, but it seems to stay inside forever. So it seems that symbols are right where Guthrie accused Tolman of leaving his rats — forever lost in thought (Guthrie, 1953). But the access relation is *not* the relation of designation to objects and relations in the world outside, although it is an essential support. Designation comes about because of two additional features. First, some of the symbols arise from transduction from the external world and initiate transduction back to the actions of the system (both to further internal processing and to the external world). The TV eyes of the NavLab van give rise via recognition systems to internal symbols and other internal symbols move the wheels, accelerator and brake shoes. But within this, the structures and the processing can be arranged so the internal structures behave according to the external environment — so that internal symbols can stand for *tree1* and *tree2*, and also the generic *tree*.

Here is another possibility. Although not directly stated, there exists a strong undercurrent in Greeno's chapter of identification of symbols with propositional expressions and what can be articulated (see especially the concluding section). Mental models are counterposed to symbolic expressions; abstract entities that have simple mappings to the external situation are seen as distinct from symbols and symbolic structures; and, in an acknowledged shadowing of Gibson, *direct* coupling with the world in normal activity is taken to bypass the need for representation altogether. This suggests that the emphasis on situated action is in part a reaction to the current

flurry of logicist interpretations of artificial intelligence (Genesereth & Nilsson, 1987). However, the theory of symbols that has arisen in computer science is certainly not tied to such an interpretation.[3]

Now, I should not get myself exercised about Feigenbaum's and Greeno's foundational interpretations — nor they about mine. Foundations are always contentious. (Sometimes it feels like that is what foundations are for.) General views of a science can be taken as heuristic. Differences in such views often lead people to explore different paths, but only occasionally do they keep them from doing good science. So, I am exceeding glad over Feigenbaum's concern with knowledge, and trust it will lead him towards getting us examples of systems with much more knowledge than we have had the courage to build to date. So, I am exceeding glad over Greeno's concern with action situated in the world, and trust it will lead him towards getting representations with the right sorts of model structure.

The Heartland

The heartland of a science is where the real work gets done in putting the science together. It is in the accumulation of a network of specific techniques for making predictions and explanations, and in our attempts at constructing an encompassing cathedral-like theory, that progress is made — or fails to be made, so leaving matters in wait until some better next generation of scientists finds a way.

More architectural issues

I need to return to the topic of architectures. My initial discussion focused on their role in putting cognitive science together, using Soar as an exemplar. But as noted in Table 2 there are lots of other candidate architectures, many on exhibit in this volume. The individualized production systems used by Larkin and by Charness are not really candidates. However, Act*/PUPS is certainly a candidate and CAPS could form the basis for one if its scope were expanded. And, although the proposal of Kosslyn and his colleagues is too nascent to be a serious candidate yet, it reminds us that the architecture sits at the boundary of the cognitive and neural bands, a place notorious for not fitting together. So, with all these architectures around, does all this fit together? Or is it just another centrifugal research area — soon to bespatter all participants, and reveal that, once more, psychology is not yet ready to get it together.

At the present time lots of architectures must exist and coexist. We do not know enough to put together a single candidate architecture — not yet. So, although I am personally attempting to develop Soar into a prime candidate, I neither wish nor recommend that work stop on others. Thus, my proposal is not that there be one unified theory embodied in an architecture, but that each and every cognitive theory should be a full-bodied architecture that can integrate results from across the breadth of cognitive phenomena. This pluralism was also a theme in my William James Lectures, accounting for the plural *theories* in its title (Newell, 1987).

The main force toward convergence will come through the successful coverage of a wide array of disparate phenomena. Of course, it is an act of scientific faith that two theories cannot explain in detail hundreds of disparate regularities across the breadth of cognition without being

---

[3]The William James Lectures discuss the relation of mental models to symbol structures and problem spaces in detail (Newell, 1987, Lecture 7; Polk & Newell, 1988).

fundamentally the same under the skin. But we need not face that eventuality until we generate it. Actually, I cannot imagine a more exciting situation than having two unified cognitive theories (how about a dozen, while we're at it), each of which makes strong quantitative predictions across perception, memory, reasoning, immediate response, knowledge acquisition, skill learning, and motor behavior — but which are also radically incommensurate. We would be the wonder of the scientific world! Even the most radical such event in scientific history — the wave-particle duality — required only a few years for a satisfactory technical synthesis to emerge, once the phenomena covered became diverse, with the quantum formulation. Of course, philosophical, heuristic and popular commentary about incommensurability of the duality continued to swirl for a much longer time. But the right place to measure the progress of science is in the living technique, not in the commentary.

In fact, the nonconvergence issue does not look to me like a major threat. Examination of the candidate architectures shows them to have an immense communality of mechanism. Act* and Soar are both built around production systems, which is to say an associational recognition system. They both work with symbolic data structures as representations, gain their directionality through goal hierarchies, and employ problem spaces as the way of formulating tasks. CAPS shares the use of symbolic structures and production systems. Some of the other aspects exist in CAPS only in rudimentary form, since it has been used mostly for the vertical integration of a single complex skill (reading), so problem spaces and goals are not really necessary in their full-blown glory.

Certainly there are differences. Table 7 shows some between Act* and Soar. Act* has both a declarative and procedural memory; Soar has only the procedural one. Act* does not have any higher levels of organization than its production system; Soar is organized in hierarchies of problem spaces, and will probably acquire additional higher organization. Act* creates its goals deliberately, by positing them as actions in productions; Soar creates its subgoals automatically by impasses. Act* controls processing by a continuous quantity (activation) which determines a variable rate of computation; Soar uses multiple production firings in the decision cycle. Act* learns by means of multiple learning mechanisms; Soar uses only chunking.

On the surface, these differences look very large. But the rate of convergence is pretty striking. Soar and Act* are the two best examples in all the world of general chunking systems. When Soar goes to take in information from the outside, which is in declarative form, its mechanism for assimilation looks like a variant of the interpretation scheme of Act* (Yost, 1988). Activation seems like a major difference. But along comes PUPS, which preserves much that is important in Act*, but abstracts away from activation. The most striking difference of all would seem to be the separate declarative memory in Act*, especially if one takes the rhetoric of Act* seriously (Anderson, 1983, p. 21). But Soar has developed ways of learning and recalling declarative data. This mechanism, *data chunking* (Rosenbloom, Laird & Newell, 1987; Rosenbloom, Laird, & Newell, 1988), is built from chunking, but constitutes a separate subsystem with its own special properties. It becomes hard to say whether Soar has one learning mechanism or two. Now, for the life of me, I don't want to say that Soar and Act* are simply the same! I do want to say that they look to me to be variant explorations of the same underlying mechanisms.

I would rather view multiple unified theories as more like an insurance policy on our getting one or two that are successful. Do you realize how much effort it will be to get a unified theory

	Act*	Soar
Memory	Declarative Procedural	Procedural
Higher Organization	None	Problem spaces
Goals	Deliberate learned	Impasse created
Control	Activation variable rate	All-or-none cycles
Learning	Compilation composition proceduralizaton Tuning strengthening discrimination generalization Declarative augmentation	Chunking

Table 7: Soar and Act*.

of cognition, with its supporting architecture and detailed explanations and predictions? At issue is not just scientific creativity — many will believe they have creative ideas for an architecture that differs from those currently in existence. The issue is person-years of efforts — hundreds of person-years to get an architecture beyond the talk stage, beyond the prototype stage, and into genuine contention. Maybe none of us will have the stamina. There are just too many phenomena out there to be covered by a unified theory.

A comprehensive architecture, such as Act*, CAPS or Soar, contains many mechanisms that have been the object of a good deal of study in cognitive science, for instance production systems and problem spaces. These architectures capture rather easily the phenomena that these mechanisms have been used to explain in other studies. However, by the same turn, these architectures do not contain other mechanisms that have played a role in cognitive research. Thus, far from putting things together, it might seem like these architectures are devices for partitioning the whole field. In some sense, this must be so. To incorporate production systems directly and not, say, schemas, is to favor one set of theoretical mechanisms over another, and thus to divide the field, at least in the short term. The situation is no worse than with any other theoretical choice, of course, but it still contrasts with my casting comprehensive architectures as ways to bring the field together.

My own solution to the tension described above is to emphasize the obligation of a candidate cognitive architecture to deal with the phenomena that important excluded mechanisms have been central to explaining. The whole point of a comprehensive architecture is to have it treat *all* the major phenomena of cognition. It is certainly the wrong turn to have the candidates partition the space of cognitive phenomena, so they talk past each other — as if emulating the worst sort of Kuhnian paradigms.

Consider, for example, that most of the architectures in this volume are built around production systems. They do not embody explicitly a notion of *schema* or *frame*. How then will they be responsive to the considerations that gave rise to these notions, and have made of schemas and frames important concepts in cognitive science?

Let us start with the evident truth that knowledge is organized. The items of knowledge relevant to the analysis of a scene or the performance of an action within a task context is strongly interrelated — they cluster around the scene or event. Furthermore, human action makes equally evident that it partakes of this organized knowledge, rapidly, effectively and in substantial quantities. It is incumbent on theories of cognition to capture this phenomena.

Schemas are a proposed solution to the imperative of the organization of knowledge in action that the human evinces. The term schema has a long and variegated history, having roots in Head's (1920) motor schemas, Bartlett's (1920) strongly memorial structures and in Piaget's (1952) action schemas. All these are highly general and diffuse theoretical constructs. The notion of schema became grounded when data structures and programs were created to capture this construct — the frames of Minsky (1975), the conceptual dependency structures of Schank (Schank & Ableson, 1977), and the schemas of Norman and Rumelhart (Norman, Rumelhart & LNR Research Group, 1975).[4] With these developments we finally obtain operational notions that proffer actual solutions.

The key feature of this operational concept of schema is positing a fixed data structure and variablizing a fixed set of places in the structure (the slots). The schema is completely rigid about the frame, while being open (or open, subject to a set of constraints) about a fixed, predetermined (i.e., rigid) set of aspects. Thus, they are devices of specific but limited adaptability.

The argument for this solution is that it holds together the in-the-large organization of related knowledge that is so evident in human behavior. This is stated clearly by Minsky:

> It seems to me that the ingredients of most theories both in artificial intelligence and in psychology have been on the whole too minute, local, and unstructured to account — either practically or phenomenologically — for the effectiveness of common sense thought. The "chunks" of reasoning, language, memory, and "perception" ought to be larger and more structured, and their factual and procedural contents must be more intimately connected in order to explain the apparent power and speed of mental activities. (Minsky, 1975, p. 211)

But this large-grain-size argument seems to me misplaced. It confuses structure with behavior. It says that if humans cluster knowledge, then the internal representation must be a pre-existing fixed structure that is that cluster. This will, of course, capture some of the action, especially in non-dynamic situations where there is no way to determine how or when the knowledge was assembled, but only that it governs the current behavior.

We need to ask how a production-based cognitive architecture is to respond to this same imperative. Such a system has a declarative representation — usually over objects defined as collections of attributes and values, where the values can themselves can be objects. This is

---

[4]Semantic nets do not quite belong to this family, and they only became so with developments such as partitioned semantic nets (Hendrix, 1977).

reminiscent of schemas and frames, though it predates them considerably (Newell & Shaw, 1957). However it has none of their characteristic additional apparatus — defaults, inheritance hierarchies, attached procedures, etc.

Collections of productions then provide the functional equivalent of complex schema or frame structures. Each production provides a link, when instantiated. Inheritance occurs by productions automatically executing on the results of others and so can march up a concept hierarchy in a context-dependent way. In Soar, for example, such a sequence occurs in a single elaboration phase (Figure 3), in an essentially automatic way. Simple attached procedures can be realized by other productions (again, in Soar, within a single decision cycle). Complex attached procedures are realized by break outs into the full power of the problem solver.

From the description I have given, it is possible to see a structure in a production system that could correspond to schemas. It has certain properties that move it in what seems the desirable direction — it is dynamic and generated on the fly in response to the local situation. Its most striking feature is its high disaggregation compared to the standard implementations. Its units are productions, which correspond to the smallest parts of the data structures of schemas and frames. We know these are the units, and not something effectively larger, from the unit of learning being the production (whether in Soar or Act*).

But all this is simply opinion, though a commentator is nothing if not a purveyor of opinions. It is meant to be an invitation to architectures that are built around production systems to address in a general and principled way how to do what schemas can do. The invitation is issued by means of a pointer to the type of phenomena where the current data-structured instantiation of a schema might reveal its limitations and where the more finely decomposed recognition systems might show a difference.

The foothills of rationality

All science has a strong tendency to work from the simple to the complex, from the more controlled to the less. However, a certain amount of scientific activity always occurs throughout the spectrum, of course, driven by interest and need. For cognitive science, there have always been arguments that complexity itself was of the essence, and even that simpler was not always easier. However, that does not gainsay the general trend. Even though the higher mental processes constituted a major focus of the earliest years of the cognitive revolution — the problem solving and organized decision-making that was Herb Simon's special concern — psychology has kept primarily to the low road of work in memory and immediate responses amenable to chronometric analysis. To give yet one more (oft noted) example: over the years research in reading has moved from the letter, to the word, to the sentence, to the paragraph (most recently), and still has ahead the page, chapter, book, encyclopedia and library.

We talk of the complexity of a task and its associated behavior, but in fact this is strongly linked to the timescale of behavior (see Table 4 again). At the scale of ~1 s, where behavior is immediate, the architecture is much in evidence. As the scale grows, more time becomes available for processing and deliberation, and the human moves toward rational behavior — which is to say, toward being characterizable by goals and the knowledge available — the world Feigenbaum declares all important.

It takes time for the human to bring to bear all that he or she knows about a problem at hand,

and it never completely happens (or mathematics would be easy). The peaks of rationality always rise up on the temporal horizon, just another ridge or two away. Much real behavior takes place on the foothills of rationality, in the range from ~10 s to ~10^4 s (a few hours). Cognitive psychology — I should say modern experimental psychology — has located itself at immediate behavior and only gradually moves up the scale. Such movement, then, becomes an indicator of putting it together. It is only possible to deal with larger time scales by bringing to bear considerations from many subareas of cognition. After all, the subject is bringing them to bear, so it stands to reason that the scientist must as well.

Table 8 shows how we are moving up to foothills of rationality. I have plotted on the timescale chart the phenomena that each contributor to the volume is primarily dealing with (by last-name initials of the authors). They cluster up at the intendedly rational band, in the minutes to ten-minutes range. This is, of course, partly a CMU speciality. But the papers from, say, the Eighth Symposium in 1973 (where the Twenty-questions commentary was given) would cluster between 1 and 10 s, two orders of magnitude lower. I take this figure as evidence of my theme that it is being put together. Of course, as befits empirical data, there are exceptions. Steve Kosslyn is focused on basic architectural issues, below 1 s, on the border rather than in the heartland. And we have had to add the historical band above 10^8 s, to take care of Herb's reflections on himself as subject.

Even as we move up the foothills, the processing limits show through in many ways. As Herb has constantly maintained, only a few parameters seem to suffice to express the effect of the architecture. These include the rate of cognitive processing, the size of STM, the size of a chunk and the rate of chunking. Observe, to pick up on an earlier theme, that this is a highly differentiated and structured set of functional parameters. It contrasts sharply with what I caricatured earlier as rationality juice, a set of anonymous and homogeneous resources or capacities. Herb's view, besides having much truth on its side, also has much to recommend it in our attempt to put it all together. For it says that a small number of constants suffice to carry out analysis over a wide range of behavior — over the whole band of intendedly rational behavior, and maybe more.

It is our task as cognitive psychologists to characterize the ways in which the underlying architecture shows through in the foothills — to find out what structures need to be used and what parameters need to be measured. Success in the venture gradually stitches together all the parts. A lot of the research presented at this conference can be seen this way. Let me just pick a couple of examples, where I have something concrete to say.

First, Kotovsky and Fallside note that their current data confirms their earlier finding (Kotovsky, Hayes & Simon, 1985), also in the Tower of Hanoi, of a long exploratory phase with a short final phase, which starts again from the beginning. They attribute this to the difficulty in performing the operators, which inhibits effective problem solving, until this is learned (during the initial phase). This of course is not the primary concern of their paper, but I find it interesting. It reminded me of some of the work in cryptarithmetic that Herb and I did (Newell & Simon, 1972). I reproduce the problem behavior graph of S3 on DONALD+GERALD in Figure 5. It will be observed that it falls into two phases, a long initial one (76%), and a short final one (24%). The subject essentially starts over one more time, and goes much further than he ever had before.

TIMESCALE OF HUMAN ACTION

Scale (secs)	Time Units	System	World (theory)
10^{10}	centuries		HISTORICAL BAND
10^{9}	decades		
10^{8}	years	S	
10^{7}	months		SOCIAL BAND
10^{6}	weeks		
10^{5}	days		
10^{4}	hours		RATIONAL BAND
10^{3}	10 mins	DK KF, A, F	
10^{2}	minutes	C, H, G, L, ES	
10^{1}	10 sec	JC	COGNITIVE BAND
10^{0}	1 sec	KSC	
10^{-1}	100 ms		
10^{-2}	10 ms		NEURAL BAND
10^{-3}	1 ms		
10^{-4}	100 μs		

Table 8: Timescale of action considered in the volume.

It is not evident that the explanation of operator difficulty works here. There is a difficulty all right, indicated by the repeated returns to one particular state. This happens to be the E+O=E column. It is indeed a puzzlement and, roughly speaking, when S3 gets it straight he is able to solve the problem. But S3 does not really seem to be learning how to apply operators in the sense of Kotovsky and Fallside.

What seems a better description of the cryptarithmetic behavior is that the subject engages in *progressive deepening*. This is a search strategy in which one passes over the same task again and again, each pass acquiring some new item of information. The strategy was first defined by DeGroot (1965) in chess. But it has much wider currency. It seems to be what is going on in Figure 5. I might conjecture that it is going on in the situations described by Kotovsky and

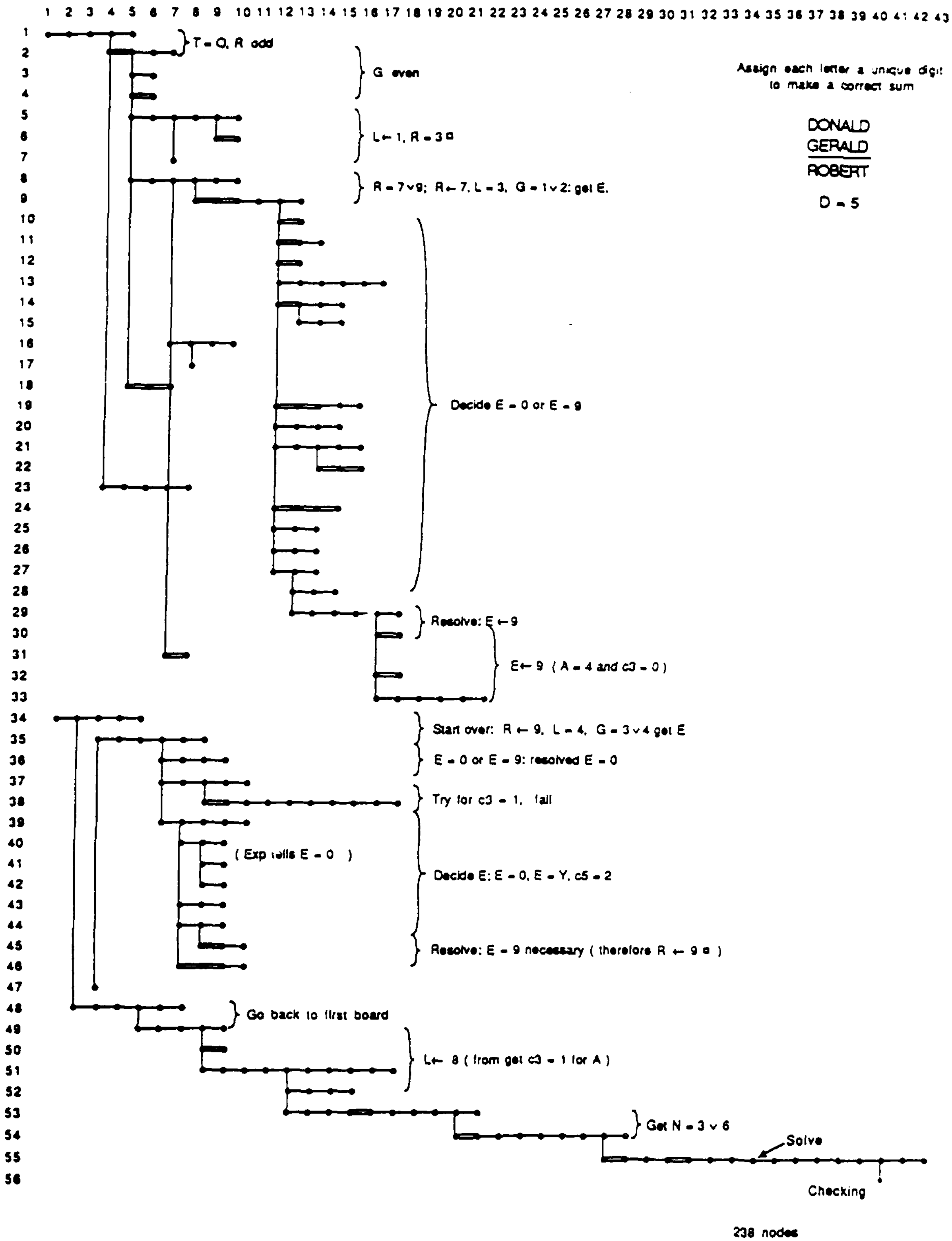


Figure 5: Progressive deepening in cryptarithmetic.

Fallside, although it is a little hard to be sure.

It might also be going on in some of the sentence construction situations that Dick Hayes describes in his chapter. He notes that subjects compose left to right adding what he calls sentence parts at the leading edge. But they do in fact repeat and correct what they do, and their behavior has some of the typical look of progressive deepening. It is not really possible to tell, because the studies are not focussed on the exact question. But the illustrative fragment of typical protocol given by Hayes looks like Figure 6 when drawn as a problem behavior graph. We can see clearly the repetitions that constitute the hallmark of progressive deepening.

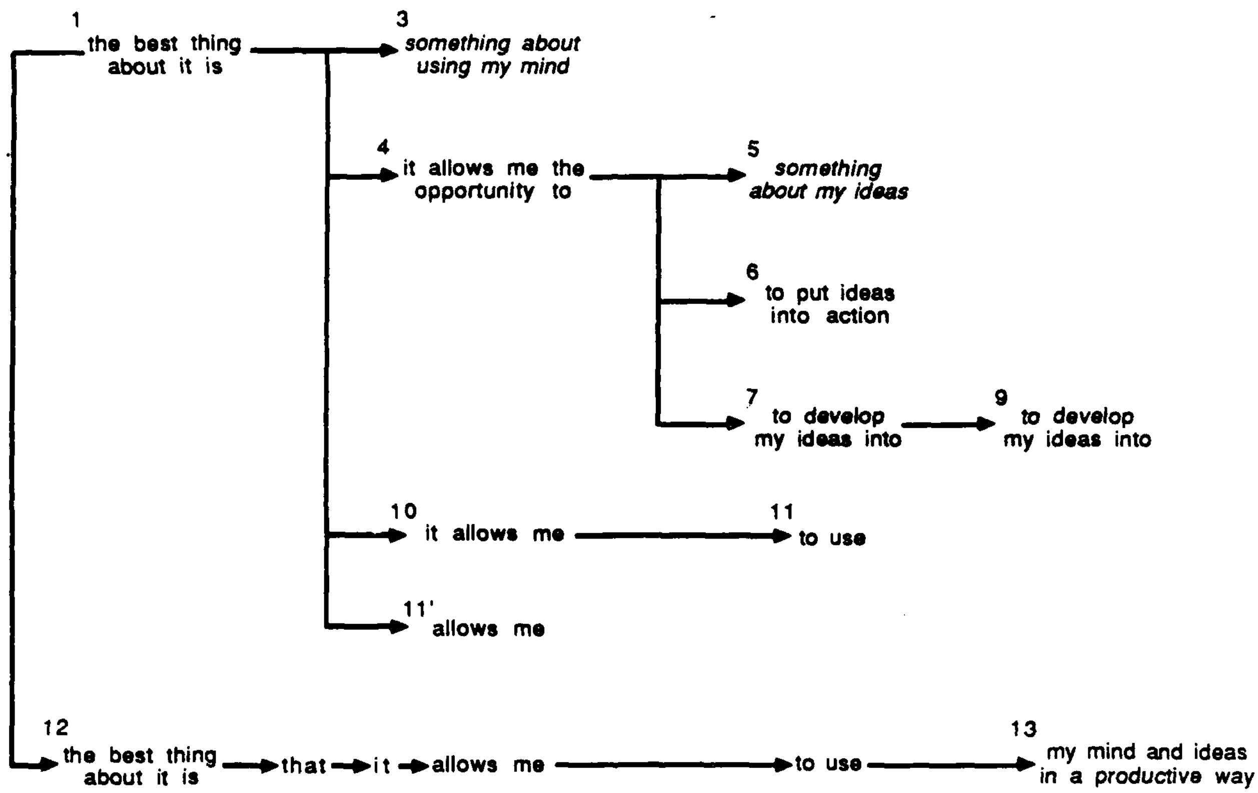


Figure 6: Problem behavior graph of sentence generation (after Hayes).

Progressive deepening seems to be even more widespread. David Steier, in attempting to construct a system for designing algorithms in Soar (Steier & Newell, 1988), has looked in some detail at protocols (Kant & Newell, 1984). What he finds is a form of progressive deepening, applied to a design task, rather than an information-gathering task. Design, of course, invariably involves successive refinement, which superficially may seem itself to be the same notion — taking successive refinements to be successive deepenings. In fact, progressive deepening provides the *control* of what to refine, by going over and ov^r again what one has done, finding the next item of relevant information (when evaluating or analyzing) or finding the next place where the design should be extended or refined.

What does this have to do with the architectural features that show through in the foothills? Progressive deepening is a pattern of behavior that appears to arise from the architecture. The

standard argument is that it is a response to memory limitations. In this regard, it is interesting to note that the method is still essentially unknown in AI, where the architectures have very different memory properties. Consequently, the ubiquity of progressive deepening is in fact a geological feature of the foothills — an example of the regularities that we need to discover. It is not clear yet that this is the case, but there are hints all around.

The second example concerns a major point of Dunbar and Klahr's chapter, which is that two problem spaces are involved in their discovery task. They give a detailed account of these two spaces, showing how they illuminate what is going on in their complex task. George Miller, in his chapter, picked up on the issue of multiple spaces, focussing on the relations of the spaces to each other and whether search could be concurrent. In fact, there are many more than two spaces involved. Table 9 gives a list, whomped up by a little arm-chair task analysis on my part.

Hypothesis space
Experimentation space
Experimental design space (program synthesis)
Experiment task environment space (predict Bigtrack)
Observation space (Bigtrack's behavior)
Data analysis space (was the experiment supported?)
Underlying mechanism space (constrain hypothesis space)
Prior literature space (what has prior theory said)

Table 9: Multiple spaces for scientific discovery in the Bigtrack world.

What does the number of problem spaces have to do with the architectural features of the foothills? I think Soar provides a clue. With Soar, we have finally found out how to have multiple problem spaces. Not one or two problem spaces, but problem spaces all the way down. Furthermore, this is driven by the architecture — scratch an impasse, get another problem space. The proliferation of spaces may be modulatable ever so slightly by deliberation, but not much. Thus, the multiple problem-space character of a task is not a strategy choice for an intelligent agent or even a task characteristic. Multiple problem spaces are a feature of the foothills, created by the nature of the cognitive architecture. That multiple spaces show up in the work of Dunbar and Klahr, or of Simon and Lea (1974) before them, is not a discovery of a specific feature of their tasks. Rather, the analysis these authors performed was thorough enough to dig out the spaces that are there. Good for them. But multiple spaces are there for all tasks in the foothills, if one just learns how to look for them.

The Borders

To put it all together requires more than just becoming unified within the heartland, though that is certainly task enough. Cognitive science has especially many borders, for the study of man is one of the great intellectual gerrymanders. Along some borders, what is required is simply to understand how things are transformed in crossing them. Along other borders, however, serious intellectual work must be done, if cognitive science is to attain any substantial unification. The contributions to the volume provide the opportunity to touch on an issue on each of two key borders, the neural band below and the social band above.

The neural band

One senses that integration along the neural border cannot be far off. This seemingly forever impenetrable border has been a feature of our scientific landscape for so long, with periodic bursts of hope and optimism, that it is difficult to know how to read the signs. Certainly, the enthusiasm for connectionism is running strong and there are additional signs as well, as Kosslyn's chapter — our lone entry from that research frontier — again attests. More certain than even the enthusiasm on the upland side of the border is the immense progress being made on the other side. Massive detail and analysis continues to build a reasonably consistent picture. The various wild alternative speculations that seemed always to be with us, seem to have finally damped down — e.g., the action is not really in the neural circuits at all, but in the glial cells, or in the macromolecules. The sense of hopelessness at not having a single viable neural candidate for the engram has given way to active investigation of a range of intriguing phenomena, such as long-term potentiation (Lynch, McGaugh & Weinberger, 1984). Neural anatomical functional specificity is way up, which makes projects such as Kosslyn's important enterprises.

The auguries for my own project of putting it all together are surely conflicting here. On the one hand, integration across this nether border cannot but help the cause. I have already noted with approval the grounding of the floating kingdom. All cognitive functions are now to be tied to an absolute time scale — recognition memory accesses at ~10 ms, elementary deliberations at ~100 ms, etc. The big implication is the potential of neural data to be brought to bear everywhere. This can only be a great thing, the ability to bring more constraints to bear (recall Table 3).

On the other hand, to hear it from the connectionists, theirs is a movement to sweep symbolic cognition away. The success of neuroscience will bring with it success at the cognitive level, but not the integration of the cognitive science that I have been extolling throughout this essay. For it is a new paradigm that will take over. It is to be a revolution a la Francais, with no ministries of the interior available. Or so they say.

I am rather partial to revolutions in the fashion of Darwin or Bohr. As I indicated earlier, revolutions within the world of paradigm science have this benign character in which, whatever the cover story, the techniques and the results continue to accumulate.

In fact, I believe that neuroscience-inspired architectures, whether of the connectionist stripe or the more functional variety of Kosslyn, are worth pursuing. It is only by such means that we will find out the implications of neural technology. The proposition that unified theories of cognition take the form of architectures holds for whatever architectural structure one considers.

I would argue that connectionist architectures should strive towards being unified theories of cognition, just like the rest of us. Good results on some special corner of phenomena is a good thing to have, and indeed is necessary. But it only means a theory has joined the microtheories, of which psychology has now quite a few. We need unified theories that cover the full gamut of psychological phenomena. Connectionists architectures must strive towards the same extended coverage that the symbolic architectures will be attaining.

The social band

The border looking upward toward social behavior has a quite different character for

individual psychology than the border looking downward toward neurophysiology. For one, a different foot wears the reductionist shoe, and when the other shoe drops, as it must in both cases, it will land on a different toe. Furthermore, both individual and social psychology belong to the same field, the really marshy boundaries being the crossovers from social psychology to sociology and anthropology. This makes the border considerably more permeable, and historically there has been lots of traffic between individual psychology and social psychology.

Still the interpenetration of cognition and the social band does not seem to me in very good shape currently. It is not for lack of trying, and by demand pull, rather than technology push, which would seem to be the right way for it to happen. But we have now had a decade of work in *social cognition*, as this attempt has come to be called. There has been no lack of enthusiasm and no lack of effort. The Seventeenth Carnegie Symposium on Cognition (Clark & Fiske, 1982) provided a mid-term report, which was pretty upbeat. But somehow it hasn't happened. A symposium now would not have much really new to say over the early 1980s. This is no place to make a real diagnosis. My carom shot, for what it is worth, is that social cognition successfully moved to variables of internal processing and internal memory. But they still kept the comparative-statics methodology. They never moved to consider *mechanisms of cognition* in the social setting — not in the way that Herb did when he went after all those topics in cognition listed in Table 1.

So psychology will have to try it again in a different way with a new set of ideas about how it might work. One of the great things about science is there is no ultimate defeat. It is the land of eternal regrouping until success is attained. My belief, consistent with my diagnosis above, is that we must deal with models of the human social actor as a system of mechanisms. But I have little faith in my own preferences here.

In thinking about the task on this border, I noted that Herb was off by himself in the timescale chart of Table 8, far away from the rest of us. This led me to list in Table 10 the fields in which Simon has made his contributions. There are lots of them, but that is not my point. Note how many are social sciences. Herb has been working across this upper boundary throughout his whole career. In fact, his most notable accomplishments, those for which he was given his Nobel Prize, are the application of the model of intendedly rational behavior to economics — one of the social sciences far off from social psychology, the object of my attention here.

Even so, Herb has not been able to abolish that border. Indeed, much of his work in these social sciences occurred early in his career. Then I went back to the lonesome outpost high up in Table 8. The reason Herb is out there is his most recent work on the psychology of scientific discovery (Langley, Simon, Bradshaw & Zytkow, 1987). That research is a curious mixture of detailed simulations of individual attainment and significance within the historical timescale. Could it be that Herb is pioneering a new way to finally get what we know about cognition into the social band?

Artificial Intelligence	(Simon, 1963)
Cognitive Psychology	(Newell & Simon ,1972)
Computer Science	(Newell, Perlis, & Simon, 1967)
Design	(Simon, 1969)
Economics	(Simon, 1979b; Simon, 1982)
Econometrics	(Ijiri & Simon, 1977)
History of Science	(Langley, Simon, Bradshaw & Zytkow, 1987)
Operations Research and Management Science	(Holt, Modigliani, Muth, & Simon, 1960)
Organization Theory	(Simon, 1957; March & Simon, 1958)
Philosophy and Foundations	(Simon, 1947; Simon & Rescher, 1966)
Philosophy of Science	(Simon, 1970)
Political Science	(Simon, 1954)
Public Administration	(Simon, Smithburg, & Thompson,1950)
Social Psychology	(Simon & Guetzkow, 1955)
Statistics	(Simon, 1955b)

Table 10: The fields of Simon's contributions (with representative citations).

Conclusion

It makes no sense to summarize a summary of something, itself a summary event. But a conclusion to a concluding essay is still conscionable. I have been upbeat in the extreme about the prospects of cognitive science being able to put it all together. I have seen in the contributions to this volume many signs that this is happening. I can think of no better way to honor Herb Simon for the immense part he has played in initiating the cognitive revolution and in giving it substance and sustenance through its first four decades, than a volume that bears witness to the maturing of cognitive science into a unified cumulating science. Such an eventuality will not surprise him, of course. Nor will he see in it any paradigm shifts, however startling the changes may seem to other observers. For it will be a continuation of the path he has been scouting for us. And it will reveal that he has been on the main path all along. That is the best present there can be for a true scientist.

References

Anderson, J. R. (1983). *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Ando, A., Fisher, F., & Simon, H. A. (1963). *Essays on the Structure of Social Science Models.* Cambridge, MA: Massachusetts Institute of Technology Press.

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review, 86,* 124-140.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Baylor, G. W., & Simon, H. A. (1966). A chess mating combinations program. *Proceedings of the 1966 Spring Joint Computer Conference.* .

Berliner, H., & Ebeling, C. (1988). Pattern knowledge and search: The SUPREM architecture. *Artificial Intelligence,* . (in press).

Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In Chase, W. G. (Ed.), *Visual Information Processing.* New York: Academic Press.

Chase, W. G. (Eds.). (1973). *Visual Information Processing.* New York: Academic.

Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In Chase, W. G. (Ed.), *Visual Information Processing.* New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Clark, M. S., & Fiske, S. T. (Eds.). (1982). *Affect and Cognition.* Hillsdale, NJ: Erlbaum.

Clark, H. H., Carpenter, P. A., & Just, M. A. (1973). On the meeting of semantics and perception. In Chase, W. G. (Ed.), *Visual Information Processing.* New York: Academic Press.

Coles, L. S. (1967). *Syntax Directed Interpretation of Natural Language.* Doctoral dissertation, Carnegie Institute of Technology,

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images.

In Chase, W. G. (Ed.), *Visual Information Processing*. New York: Academic Press.

De Groot, A. (1965). *Thought and Choice in Chess*. The Hague: Mouton.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Feldman, J. A., & Ballard, D. (1982). Connectionist models and their properties. *Cognitive Science, 6*, 205-254.

Fisher, D. F., Monty, R. A., & Senders, J. W. (Eds.). (1981). *Eye, Movements: Cognition and Visual Perception*. Hillsdale, NJ: Erlbaum.

Forgy, C. L. (July 1981). *OPS5 User's Manual* (Tech. Rep.). Computer Science Department, Carnegie-Mellon University,

Genesereth, M. R., & Nilsson, N. J. (1987). *Logical Foundations of Artificial Intelligence*. Los Altos, CA: Morgan Kaufman.

Gilmartin, K. J., Newell, A., & Simon, H. A. (1976). A program modeling short-term memory under strategy control. In Cofer, C. M. (Ed.), *The structure of human memory*. San Francisco, CA: W. H. Freeman & Company.

Gregg, L. W., & Simon, H. A. (1967). Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology, 4*, 246-276.

Groner, R., Menz, C., Fisher, D. F., Monty, R. A. (Eds.). (1983). *Eye, Movements and Psychological Functions: International views*. Hillsdale, NJ: Erlbaum.

Guthrie, E. R. (1953). *The Psychology of Learning*. New York: Harper and Row.

Head, H. (1920). *Studies in Neurology*. Oxford, England: Oxford University Press.

Hendrix, G. G. (1977). Expanding the utility of semantic networks through partitioning. *Proceedings of the 5th International Joint Conference on Artificial Intelligence*. Menlo Park, CA, AAAI.

Holt, C. C., Modigliani, F., Muth, J., & Simon, H. A. (1960). *Planning Production, Inventories, and Work Force*. Englewood Cliffs, NJ: Prentice-Hall.

Ijiri, Y., & Simon, H. A. (1977). *Skew Distributions and the Sizes of Business Firms*.

Amsterdam, The Netherlands: North-Holland Publishing Company.

Kant, E. & Newell, A. (1984). Problem solving techniques for the design of algorithms. *Information Processing and Management, 20*, 97-118.

Klahr, D. (1973). Quantification processes. In Chase, W. G. (Ed.), *Visual Information Processing*. New York: Academic Press.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problem hard? Evidence from the Tower of Hanoi. *Cognitive Psychology, 17*, 248-294.

Kuhn. T. (1962). *Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Kulkarni, D., & Simon, H. A. (1988). Processes of scientific discovery: The strategy of experimentation. *Cognitive Science, 12*, 139-175.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*(1), 1-64.

Laird, J. E., Rosenbloom, P. S., & Newell, A. (1986). Chunking in Soar: The anatomy of a general learning mechanism. *Machine Learning, 1*, 11-46.

Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific Discovery: Computational explorations of the creative processes*. Cambridge, MA: MIT Press.

Larkin, J. H., & Simon, H. A. (1987). What a diagram is (sometimes) worth 10,000 words. *Cognitive Science, 11*, 65-100.

Larkin, J. H., McDermott, J., Simon, D., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 1335-1342.

Lynch, G., McGaugh, J. L., & Weinberger, N. M. (Eds.). (1984). *Neurobiology of Learning and Memory*. New York: Guilford Press.

March, J. G., & Simon, H. A. (1958). *Organizations*. New York, NY: Wiley.

Minsky, M. (1975). A framework for the representation of knowledge. In Winston, P. (Ed.), *The Psychlogy of Computer Vision*. New York: McGraw-Hill.

Monty, R. A., & Senders, J. W. (Eds.). (1976). *Eye Movements and Psychological Processes*. Hillsdale, NJ: Erlbaum.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Neves, D. M. (1981). *Learning Procedures from Examples*. Doctoral dissertation, Carnegie Mellon University,

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In Chase, W. G. (Ed.), *Visual Information Processing*. New York: Academic Press.

Newell, A. (1973). Production systems: Models of control structures. In Chase, W. C. (Ed.), *Visual Information Processing*. New York: Academic Press.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*, 135-183.

Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18*, 87-127.

Newell, A. (1986). Scale counts in cognition. APA Scientific Contribution Award Lecture, American Psychology Association Convention.

Newell, A. (1987). Unified Theories of Cognition. The William James Lectures. Harvard University, Spring 1987. (Available in videocassette, Psychology Department, Harvard).

Newell, A. & Shaw, J. C. (1957). Programming the Logic Theory Machine. *Proceedings of the 1957 Western Joint Computer Conference*. .

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newell, A. & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19*(3), 113-126.

Newell, A., Perlis, A. J., & Simon, H. A. (1967). What is computer science? *Science, 155*, 1373-1374. (letter).

Newell, A., Shaw, J. C. & Simon, H. A. (1957). Empirical explorations of the Logic Theory Machine: A case study in heuristics. *Proceedings of the 1957 Western Joint Computer Conference*. Western Joint Computer Conference.

Norman, D. A., Rumelhart, D. E., & LNR Research Group. (1975). *Explorations in Cognition*. New York: Freeman.

Paige, J. M., & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In Kleinmuntz, B. (Ed.), *Problem Solving*. New York, NY: John Wiley & Sons.

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: International Universities Press.

Polk, T. A., & Newell, A. (1988). Modeling human syllogistic reasoning in Soar. *Proceedings Cognitive Science Annual Conference — 1988*. Cognitive Science Society, (in press).

Posner, M. I. (1973). Coordination of internal codes. In Chase, W. G. (Ed.), *Visual Information Processing*. New York: Academic Press.

Quillian, R. (1967). *Semantic Memory*. Doctoral dissertation, Carnegie Institute of Technology,

Rosenbloom, P. S., Laird, J. E., & Newell, A. (1987). Knowledge-level learning in Soar. *Proceedings of the American Association of Artificial Intelligence - 1987*. Los Altos, CA, Morgan Kaufman.

Rosenbloom, P. S., Laird, J. E., & Newell, A. (1988). The chunking of skill and knowledge. In Bouma, H. & Elsendoorn, B. A. G. (Ed.), *Working Models of Human Perception*. London: Academic Press. (in press).

Schank, R. & Ableson, R. (1977). *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Senders, J. W., Fisher, D. F., & Monty, R. A. (Eds.). (1978). *Eye Movements and the Higher Psychological Functions*. Hillsdale, NJ: Erlbaum.

Shaw, J. C., Newell, A., Simon, H. A., & Ellis, T. O. (1959). A command structure for complex information processing. *Proceedings of the Western Joint Computer Conference*. .

Simon, H. A. (1947). The axioms of Newtonian mechanics. *The Philosophical Magazine, 38*, 888-905.

Simon, H. A. (1954). Bandwagon and underdog effects and the possibility of election predictions. *Public Opinion Quarterly, 18*, 245-253.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99-118.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika, 42*, 425-440.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Simon, H. A. (1957). *Administrative Behavior, 2nd Edition.* New York, NY: Macmillan.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 26*, 467-482.

Simon, H. A. (1963). Experiments with a heuristic compiler. *Journal of the Association for Computing Machinery, 10*, 493-506.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review, 74*, 29-39.

Simon, H. A. (1969). The sciences of the artificial. Cambridge Massachusetts Institute of Technology Press. (Karl Taylor Compton Lectures).

Simon, H. A. (1970). The axiomatization of physical theories. *Philosophy of Science, 37*, 16-26.

Simon, H. A. (1974). How big is a chunk? *Science, 183*, 482-488.

Simon, H. A. (1979). *Models of Thought.* New Haven, CT: Yale University Press.

Simon, H. A. (1979). Rational decision making in business organizations. *American Economic Review, 69*, 493-513. (Nobel Lecture).

Simon, H. A. (1980). How to play twenty questions with nature and win. In Cole, R. (Ed.), *Perception and Production of Fluent Speech.* Hillsdale, NJ: Erlbaum.

Simon, H. A. (1980). Cognitive science: The newest science of the artificial. *Cognitive Science, 4*, 33-46.

Simon, H. A. (1982). *Models of Bounded Rationality.* Cambridge, MA: The Massachusetts Institute of Technology Press. (2 volumes).

Simon, H. A., & Barenfeld, M. (1969). Information-processing analysis of perceptual processes in problem solving. *Psychological Review, 76*, 473-483.

Simon, H. A., & Feigenbaum, E. A. (1964). An information-processing theory of some effects of similarity, familiarization, and meaningfulness in verbal learning. *Journal of Verbal*

Learning and Verbal Behavior, 3, 385-396.

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology, 5*, 29-46.

Simon, H. A., & Guetzkow, H. (1955). A model of short- and long-run mechanisms involved in pressures toward uniformity in groups. *Psychological Review, 62*, 56-68.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology, 8*, 165-190.

Simon, H. A., & Kadane, J. B. (1975). Optimal problem-solving search: All-or-none solutions. *Artificial Intelligence, 6*, 235-247.

Simon, H. A., & Kotovsky, K. (1963). Human acquisiiton of concepts for sequential patterns. *Psychological Review, 70*, 534-546.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In Gregg, L. W. (Ed.), *Knowledge and Cognition*. Potomac, MD: Lawrence Erlbaum Associates.

Simon, H. A., & Rescher, N. (1966). Cause and counterfactual. *Philosophy of Science, 33*, 323-340.

Simon, H. A., & Siklossy, L. (Eds.). (1972). *Representation and Meaning: Experiments with information processing systems*. Englewood Cliffs, NJ: Prentice Hall.

Simon, H. A., Smithburg, D. W., & Thompson, V. A. (1950). *Public Administration*. New York, NY: Alfred A. Knopf.

Steier, D., & Newell, A. (1988). Integrating Multiple Sources of Knowledge into Designer-Soar, an Automatic Algorithm Designer. *Proceedings American Association of Artificial Intelligence — 1988*. American Association of Artificial Intelligence, (in press).

Steier, D. E., Laird, J. E., Newell, A., Rosenbloom, P. S., Flynn, R. A., Golding, A., Polk, T. A., Shivers, O. G., Unruh, A. & Yost, G. R. (1987). Varieties of Learning in Soar: 1987. In *Proceedings of the Fourth International Workshop on Machine Learning*. Los Altos, CA: Morgan Kaufman.

Thibadeau, R., Just, M. A., & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science, 6*, 157-203.

Thorp, C. E., Herbert, M., Kanade, T., & Shafer, S. (1987). Vision and navigation for the Carnegie-Mellon Navlab. Traub, J. (Ed.), *Annual Review of Computer Science*. .

Williams, T. G. (1965). *Some Studies in Game Playing with a Digital Computer*. Doctoral dissertation, Carnegie Institute of Technology,

Williams, D. S. (1969). *Computer Program Organization Induced by Problem Examples*. Doctoral dissertation, Carnegie Mellon University,

Yost, G. (March 1988). *TAQ Reference Manual* . Soar Project, Computer Science Department, Carnegie-Mellon University,

Zhang, G., & Simon, H. A. (1985). STM capacity chinese words and idioms: Chunking and the acoustical loop hypothesis. *Memory and Cognition, 13*, 193-201.